

**UNSUPERVISED BUILDING AND EXPLOITATION OF COMPOSITE
DESCRIPTORS**

Abstract of the Disclosure

5 Generally, the present invention provides a way of determining in an
unsupervised manner additional members for a family that is defined initially through
exemplar sequences. The present invention is unsupervised in that it proceeds without
any information related to the exemplar sequences defining the family, without aligning
the sequences, without prior knowledge of any patterns in the exemplar sequences, and
10 without knowledge of the cardinality or characteristics of any features that may be present
in the exemplar sequences. In one aspect of the invention, a method is used to take a set
of unaligned sequences and discover several or many patterns common to some or all of
the sequences. These patterns can then be used to determine if candidate sequences are
members of the family. In another aspect of the invention, a method is used to take a set
15 of sequences and to determine a set of maximal patterns common to a number of
sequences. The maximal patterns are determined without any previous knowledge about
any properties or features that may be present in the processed sequences.

20 1500-148.APP